

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-256000

(43)Date of publication of application : 21.09.2001

(51)Int.Cl.

G06F 3/06

(21)Application number : 2000-064296

(71)Applicant : NEC ENG LTD

(22)Date of filing : 09.03.2000

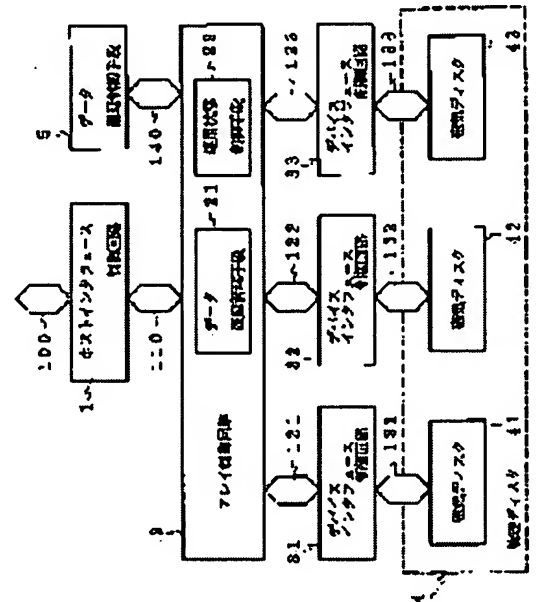
(72)Inventor : TSURUMAKI MASAYOSHI

(54) DISK ARRAY DEVICE AND DEGENERATION CONTROL METHOD USED FOR THE SAME

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a highly reliable disk array device by holding the high velocity of an I/O response and data quantity within a fixed time, and reducing the degeneracy rate.

SOLUTION: Time information required for data transfer and degenerate instruction information is transmitted from an array control circuit Via an operating state control bus 140 to a data delay controlling means 5. Also, the data delay control means 5 is provided with a function for specifying a magnetic disk, in which the data transfer delay is continuously generated from the time information required for the data transfer for each I/O of all magnetic disks 41-43 which constitute a logical disk 4 which is supplied from a data delay managing means 21 of the array control circuit 2 and a function for supplying instructions to separate the specified magnetic disk.



## LEGAL STATUS

[Date of request for examination] 26.12.2001

[Date of sending the examiner's decision of rejection] 22.07.2003

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

\* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

DETAILED DESCRIPTION

---

[Detailed Description of the Invention]

[0001]

[Field of the Invention] Especially this invention relates to the disk array equipment currently widely used as data storage with the information processor about disk array equipment.

[0002]

[Description of the Prior Art] Conventionally, when the I/O (data transfer) delay by the abnormalities in a medium of one set of a magnetic disk etc. occurs in disk array equipment, the magnetic disk which abnormalities generated is temporarily separated from a logical disk, and either of the approaches which performs the perfect separation which requires exchange of the approach of preventing the I/O process as disk array equipment being overdue or a magnetic disk is taken.

[0003]

[Problem(s) to be Solved by the Invention] In the case of the approach of separating a magnetic disk temporarily with the conventional disk array equipment mentioned above, when I/O delay occurs If abnormalities occur to one magnetic disk of the logical disks and I/O delay occurs again, after building into a logical disk the magnetic disk separated temporarily (by the high order host side) When it seems that multiple-times I/O delay occurred within a certain fixed time amount, it will become impossible to transmit the amount of data which host system needs in a certain time amount.

[0004] moreover, in the case of the approach of performing the perfect separation (it changing to a degenerate state) which requires exchange of a magnetic disk Although I/O delay is in allowance of a high order host since the magnetic disk was detached without consideration of the amount of data needed in time amount with a high order host that is, Since degeneration control was carried out in the processing time of single I/O by the side of a disk array, the rate of degeneration will become large and will make dependability of disk array equipment low.

[0005] Then, it is in the object of this invention offering the disk array equipment which can cancel the above-mentioned trouble, can aim at reservation of the amount of data within the rapidity of an I/O response, and a certain fixed time amount, can make the rate of degeneration low, and can make dependability high.

[0006]

[Means for Solving the Problem] The logical disk with which the disk array equipment by this invention consists of two or more magnetic disks, A monitor means to be disk array equipment including the array control circuit which controls the employment condition in said two or more magnetic disks, and to supervise the data delay in the data transfer between said two or more magnetic disks, It has a means to specify the magnetic disk which data transfer delay generates continuously from the monitor result of said monitor means, and a means to publish the separation instruction to the specified magnetic disk.

[0007] The logical disk with which the degeneration control approach by this invention consists of two or more magnetic disks, The step which is the degeneration control approach of disk array equipment including the array control circuit which controls the employment condition in said two or more magnetic disks, and supervises the data delay in the data transfer between said two or more magnetic disks, It has the step which specifies the magnetic disk which data transfer delay generates continuously from the monitor result, and the step which publishes the separation instruction to the specified magnetic disk.

[0008] Namely, the host interface by which the disk array equipment of this invention is connected to a host computer, The host interface control circuit which controls a host interface, The array control circuit equipped with a data delay monitor means to supervise the employment state control means and data delay which control the employment condition in the magnetic disk connected to the host bus connected to a host interface control circuit, and a host bus, The array bus of N individual (N is two or more integers) connected to an array control circuit, The device interface control circuit of N individual connected to each of the array bus of N individual, The device interface bus of N

individual connected to the device interface control circuits of each of N individual, In the disk array equipment containing M sets (M is an integer more than N) of the magnetic disks connected with the device interface bus of N individual The function for specifying the magnetic disk which data transfer delay generates continuously from the hour entry concerning the data transfer for every I/O of all the magnetic disks that constitute the logical disk supplied from the data delay monitor means of an array control circuit, and its specified separation instruction of a magnetic disk It has the data delay control means with the function to supply.

[0009] It becomes possible to separate only the magnetic disk which affects a high order host with the disk array equipment of this invention by constituting as mentioned above, and it becomes possible to separate only the magnetic disk which data transfer delay generates continuously. That is, with the disk array equipment of this invention, when data transfer delay occurs continuously, it becomes possible to separate the magnetic disk which causes data transfer delay by the above-mentioned data delay control means.

[0010] In addition, in the system of a VOD (Video On Demand) application, the data for for several seconds (for 3 - 5 seconds) are stored in the buffer by the high order host between systems operation, and even when I/O delay occurs in single shot, it is designed so that there may be no effect in a system. Although disk array equipment which can separate only the magnetic disk which data transfer delay generates continuously, can make the rate of degeneration low, and can make dependability high is desired in the system of this VOD application, offer of the disk array equipment suitable for such a system is attained in this invention.

[0011]

[Embodiment of the Invention] Next, one example of this invention is explained with reference to a drawing. Drawing 1 is the block diagram showing the configuration of the disk array equipment by one example of this invention. In drawing 1, the disk array equipment by one example of this invention consists of the host interface control circuit 1, the array control circuit 2, three device interface control circuits 31-33, logical disks 4, and data delay control means 5, the array control circuit 2 was equipped with the data delay management tool 21 and the employment state control means 22, and the logical disk 4 is equipped with three sets of magnetic disks 41-43.

[0012] It connects with the host computer which is not illustrated through a host interface 100, and the host interface control circuit 1 controls the host interface 100. It connects with the host interface control circuit 1 through the host bus 110, and the array control circuit 2 is connected to three device interface control circuits 31-33 through three array buses 121-123. Moreover, in the array control circuit 2, the data delay in the data transfer between magnetic disks 41-43 is supervised, and the employment condition in magnetic disks 41-43 is controlled by the data delay management tool 21 with the employment state control means 22.

[0013] Three 31 to device interface control circuit 33 each is connected to three sets of magnetic disks 41-43 through three device interface buses 131-133.

[0014] It is connected with the array control circuit 2 through the employment state-control bus 140, and a data delay control means 5 has the function (issuance) which supplies the function for specifying the magnetic disk which data-transfer delay generates continuously from the hour entry concerning the data transfer for every I/O of all the magnetic disks 41-43 that constitute the logical disk 4 supplied from the data delay management tool 21 of the array control circuit 2, and its specified separation instruction of a magnetic disk. The hour entry which started data transfer from the array control circuit 2 through the employment state control bus 140, and degeneration directions information are transmitted to the data delay control means 5.

[0015] In the disk array equipment by one example of this invention, magnetic disks 41 and 42 are used as a data disk, a logical disk [LUN] 4 is set up by using a magnetic disk 43 as a redundancy disk, and level 3 and an employment condition show [ RAID level ] the all seems well, respectively.

[0016] When a read-out instruction is published to disk array equipment, the host interface control circuit 1 receives this instruction, and transmits it to the array control circuit 2. The array control circuit 2 determines the magnetic disk which performs read-out-processing with reference to configuration information from the employment state control means 22.

[0017] Next, the array control circuit 2 publishes a read-out processing instruction to the device interface control circuits 31-33 through the array buses 121-123. The device interface control circuits 31-33 will publish a read-out instruction to magnetic disks 41-43 through the device buses 131-133 connected to each, if a read-out processing instruction is received.

[0018] At this time, the data delay management tool 21 of the array control circuit 2 operates the timer (not shown) assigned to each magnetic disk 41-43, and supervises the number of data transfer. Magnetic disks 41-43 perform read-out processing (data transfer) with the device interface control circuits 31-33 through the device buses 131-133 connected to each.

[0019] Then, the device interface control circuits 31-33 transmit each read-out data and a read-out processing activation result to the array control circuit 2 through the array buses 121-123. The array control circuit 2 judges normality with

reference to the read-out data and the read-out processing activation result which have been transmitted through the device interface control circuits 31-33 and the array buses 121-123 from device interfaces 131-133. Moreover, it judges whether the array control circuit 2 should carry out a state transition from a read-out processing activation result.

[0020] In addition, the array control circuit 2 performs a parity check in the data from 41 to magnetic-disk 43 each, and when parity is normal, it transmits normal read-out data to a host computer at any time through the host bus 110, the host interface control circuit 1, and a host interface 100.

[0021] Moreover, after termination of data transfer, if a read-out processing activation result is transmitted, the array control circuit 2 will perform read-out processing result transfer processing, and will end a read-out instruction. The array control circuit 2 will not be different from the disk array equipment of a Prior art at all, if normal read-out processing is performed.

[0022] Drawing 2 is drawing showing an example of the accumulation timer value table managed according to the magnetic disk used for the data delay control means 5 of drawing 1. In drawing 2, the accumulation timer value is stored in the accumulation timer value table of the data delay control means 5 every magnetic disk 41-43.

[0023] Drawing 3 is drawing showing the relation of the accumulation timer value of a magnetic disk and time amount which data transfer delay generated in one example of this invention, and drawing 4 is a flow chart which shows actuation when the data transfer delay in one example of this invention occurs. Actuation when the failure which becomes a magnetic disk 41 with data transfer delay with reference to these drawing 1 - drawing 4 occurs is explained.

[0024] When the failure which becomes a magnetic disk 41 with data transfer delay of the abnormalities in a medium etc. occurs, the data delay control means 5 adds "the time amount (for example, 1.5 seconds) which the magnetic disk 41 spent on the I/O process" sent to the storage area applicable to the magnetic disk 41 of the accumulation timer value table assigned to each magnetic disk 41-43 through the employment state control bus 140 from the data delay management tool 21. Here, the initial value of an accumulation timer value is 0 second.

[0025] Whenever data transfer delay occurs, the data delay control means 5 repeats the above-mentioned processing, and performs it, and when it becomes the value to which accumulated exceeded the threshold (for example, 3.0 seconds), the separation instruction of a magnetic disk 41 is reported to the array control circuit 2 equipped with the employment state control means 22 through the employment state control bus 140. The relation between a time-axis and the amount of data transfer is shown for the storage area of the accumulation timer value assigned to each magnetic disk 41-43 at drawing 2 in drawing 3.

[0026] the magnetic disks 41-43 shown in drawing 2 -- each accumulated -- magnetic disks 41-43 -- the value in alignment with the time-axis shown in drawing 3 which came out, respectively and was managed is stored. Between 0-A shown in drawing 3, 1.5 seconds which the magnetic disk 41 spent on the I/O process are added. It subtracts between A-B at a rate of the amount of data transfer per second of the amount of data processing / this disk array equipment per second of host system.

[0027] Between B-C, 1.5 seconds which the magnetic disk 41 spent on the I/O process are added like between 0-A. Between C-D, it subtracts like between A-B at a rate of the amount of data transfer per second of the amount of data processing / this disk array equipment per second of host system. Between D-E, like between 0-A, 1.5 seconds spent on the I/O process are added, and the magnetic disk 41 is in the condition of one-step this side which empties the data for 3.5 seconds which were storing host system.

[0028] A separation instruction is published in the place which was below remaining 0.5 seconds of the amount of data in this disk array equipment. The array control circuit 2 separates a magnetic disk 41 with this report, and it carries out a state transition to a degenerate state. By separating the magnetic disk 41 with the cause of data transfer delay, degree I/O can be resumed without delay of data transfer.

[0029] The disk array equipment of this invention securing the rapidity of an I/O response as the above-mentioned explanation, the amount of data in a certain fixed time amount which host system requires can be supplied, the rate of degeneration becomes low, and reliable disk array equipment can be offered.

[0030] If the data delay management tool 21 of the array control circuit 2 is judged to be those with a data delay disk (drawing 4 step S1), it is set to time delay =Y of a magnetic disk 41 (drawing 4 step S2), and sends out time delay =Y of the magnetic disk 41 to the data delay control means 5 through the employment state control bus 140.

[0031] The data delay control means 5 adds time delay =Y of the magnetic disk 41 sent to the accumulation timer value X of the storage area applicable to the magnetic disk 41 of an accumulation timer value table through the employment state control bus 140 from the data delay management tool 21 (drawing 4 step S3). ( $X=X+Y$ )

[0032] The data delay control means 5 reports the separation instruction of (drawing 4 step S4) and a magnetic disk 41 to the employment state control means 22 of the array control circuit 2 through the employment state control bus 140, when the accumulation timer value X which added time delay =Y of a magnetic disk 41 exceeds a threshold ( $X \geq 3.0S$ ) (drawing 4 step S5).

[0033] The employment state control means 22 separates a magnetic disk 41 with this report, and it carries out a state

transition to a degenerate state ( drawing 4 step S6). By separating the magnetic disk 41 with the cause of data transfer delay, degree I/O can be resumed without data transfer delay.

[0034] In addition, when the rate ratio of the amount of data around the second of host system and the amount of data around the second of disk array equipment which are not illustrated to the data delay control means 5 is set to Z, "X=X-Z" is notified by interruption from host system and the accumulation timer value X is stored in an accumulation timer value table.

[0035] Thus, by supervising the data transfer condition (the number of data transfer) of a magnetic disk 41 - 43 each in a small time basis (for example, 1.0S), control which separates the magnetic disk of abnormalities temporarily can be performed, and an I/O response can be made high-speed.

[0036] Moreover, the magnetic disk (magnetic disk leading to data transfer delay) which abnormalities generated is detected. It manages the amount of data of which was able to be transmitted to host system in fixed time amount with the magnetic disk. Fixed magnitude with total of the processing time of two or more I/O which should be processed within a certain fixed time amount which host system has managed The rate of degeneration can be made low by what the magnetic disk of relevance is separated for only when exceeding (3.0S [ for example, ]) (an array employment condition is shifted to degeneration) (when the amount of data which host system needs becomes close to 0).

[0037] By this control, since reservation of the amount of data to host system can also be performed, for a VOD application, reservation of the amount of data within the rapidity of the I/O response made especially important and a certain fixed time amount can be aimed at, the rate of degeneration becomes low, and reliable disk array equipment can be offered.

[0038]

[Effect of the Invention] In the disk array equipment which includes the logical disk which consists of two or more magnetic disks, and the array control circuit which controls the employment condition in two or more magnetic disks according to this invention as explained above The data delay in the data transfer between two or more magnetic disks is supervised. By specifying the magnetic disk which data transfer delay generates continuously from the monitor result, and publishing the separation instruction to the specified magnetic disk Reservation of the amount of data within the rapidity of a data transfer response and a certain fixed time amount can be aimed at, and it is effective in the ability to make the rate of degeneration low and make dependability high.

---

[Translation done.]